RTLLM: An Open-Source Benchmark for Design RTL Generation with Large Language Model

Abstract

Inspired by the recent success of large language models (LLMs) like ChatGPT, researchers start to explore the adoption of LLMs for agile hardware design, such as generating design RTL based on natural-language instructions. However, in existing works, their target designs are all relatively simple and in a small scale, and proposed by the authors themselves, making a fair comparison among different LLM solutions challenging. In addition, many prior works only focus on the design correctness, without evaluating the design qualities of generated design RTL. In this work, we propose an open-source benchmark named RTLLM, for generating design RTL with natural language instructions. To systematically evaluate the auto-generated design RTL, we summarized three progressive goals, named syntax goal, functionality goal, and design quality goal. This benchmark can automatically provide a quantitative evaluation of any given LLM-based solution. Furthermore, we propose an easy-to-use yet surprisingly effective prompt engineering technique named self-planning, which proves to significantly boost the performance of GPT-3.5 in our proposed benchmark.